RNAi is a critical determinant of centromere evolution in closely related fungi

Vikas Yadav, Sheng Sun, R. Blake Billmyre, Bhagya C. Thimmappa, Terrance Shea, Robert Lintner, Guus Bakkeren, Christina A. Cuomo, Joseph Heitman and Kaustuv Sanyal

Supplementary Information

Supplementary Materials and Methods

Construction of mCherry-tagged strains

The strategy for constructing strains expressing an N-terminal epitope tagged CENP-A and a C-terminal tagged CENP-C in the *C. neoformans* background was described previously (1). While the mCherry-CENP-A fusion was expressed ectopically from its own promoter, CENP-C-mCherry was expressed from the endogenous locus. We used the mCherry-CENP-A cassette constructed for *C. neoformans* to transform *C. deneoformans* and *C. deuterogattii* to study CENP-A localization patterns in these species. For tagging of CENP-C with mCherry at its C-terminal in both *C. deneoformans* and *C. deuterogattii*, the constructs were generated by overlap PCR as described previously (2). Briefly, about 1 kb region of the gene sequence upstream to the stop codon (US) (using primers VYP501-502 and VYP701-702) and another 1 kb sequence downstream (DS) (using primers VYP505-506 and VYP705-706) of the stop codon was amplified from the genome of each species. A 3.2 kb long sequence fragment containing mCherry along with the neomycin (mCh-Neo) gene was amplified from a plasmid, pLKB25 (using primers VYP503-504 and VYP703-704) (2). These three amplified DNA fragments (US, mCh-Neo, and DS) were purified separately and then mixed in an equimolar ratio. The mix was used as a template for the final overlap PCR using primers VYP507-508 and VYP707-708. The overlap product of approximately 5.2 kb was used to transform *C. deneoformans* or *C. deuterogattii* using biolistics as described previously (3). The transformants were selected on YPD medium containing 200 µg/ml of

G418 (Sigma-Aldrich). The transformants were screened by PCR to confirm the integration of mCherry-encoding sequence at the 3' end of the target gene. The tagged strains were then imaged using a DeltaVision (GE Healthcare) microscope. The images were processed using ImageJ and Adobe Photoshop.

Chromatin-immunoprecipitation

Chromatin immunoprecipitation (ChIP) assays were carried out as described previously (4). Briefly, mCherry tagged CENP-A or CENP-C strains were grown in 100 ml YPD to OD_{600} = 1. Formaldehyde was added as the cross-linker to a final concentration of 1%, and the mix was kept at room temperature for 30 min with intermittent shaking. The fixed cells were harvested and resuspended in 10 ml of water containing 0.5 ml of 2-Mercaptoethanol (Sigma-Aldrich). The cell suspension was incubated at 30°C for 1 h followed by spheroplasting using the lysing enzyme from *Trichoderma harzianum* (Sigma-Aldrich). Spheroplasts were resuspended in 1 ml of lysis buffer (50 mM HEPES, pH 7.5/140 mM NaCl/1 mM EDTA/0.1% Na-deoxycholate/1% Triton-X), sonicated to shear chromatin using a Bioruptor (Diagenode) for 24 cycles of 15 s on and 15 s off bursts at the high level, and fragmented chromatin was isolated by centrifugation. The average chromatin fragment sizes ranged from 300 to 500 bp. A part of the chromatin fraction (100 µl i.e. $1/10^{th}$ volume) was kept for input DNA (I) preparation and the remaining chromatin solution was divided into two halves (450 µl each). In one of the tubes, 20 µl of RFP-TRAP beads (ChromoTek) were added and used as IP DNA with antibodies (+). In another tube, 20 µl of blocked agarose beads (ChromoTek) were added to serve as a negative control (-). The tubes were incubated at 4°C for 8 h on a rotator. The beads were then washed, and bound chromatin was eluted in 500 µl of elution buffer (1% SDS/0.1M $NaHCO_3$). All three fractions (I, + and -), were decrosslinked and DNA was isolated using phenol: chloroform extraction followed by

ethanol precipitation. The precipitated DNA was air dried and dissolved in 25 µl of MilliQ water containing 25 µg/ml RNase (Sigma-Aldrich). I and + samples were subjected to ChIP-sequencing (mCherry-CENP-A and CENP-C-mCherry in *C. neoformans*, CENP-C-mCherry in *C. deuterogattii*) to identify centromere regions across the genome. All three samples (I, + and -) of CENP-C ChIP were subjected to qPCR with centromere-specific primers along with a non-centromeric primer set. The fold enrichment for the same was calculated and plotted using GraphPad Prism.

C. neoformans and C. deuterogattii PacBio sequencing and assembly update

The *C. neoformans* (H99) and *C. deuterogattii* (R265) genomes were sequenced using PacBio sequencing to improve sequence sequence assembly of the centromeric regions. PacBio filtered subreads were used for a higher order scaffolding using SSPACE-LongRead v1-1 (5), requiring 5 linking reads (-l 5) and a 200 base gap between scaffolds (-g 200). The *de novo* assembly of the PacBio reads led to generation of 20 and 27 scaffolds for *C. neoformans* and *C. deuterogattii*, respectively. The centromere flanking gene sequences from the available GenBank assembly for *C. neoformans* (GCA_000149245.3) and *C. deuterogattii* (GCA_000149475.3) were searched using the BLAST analysis against the newly assembled PacBio assembly to identify the centromere locations. This analysis led to mapping of 10 centromeres (out of 14) in the newly assembled *C. neoformans* genome with good read depth and no sequence gaps. These completely assembled 10 centromeres are *CEN1, CEN2, CEN4, CEN6, CEN7, CEN8, CEN9, CEN10, CEN12*, and *CEN13*. One of the four centromeres that remained incomplete even after using PacBio reads was *CEN5* with two sequence gaps. Using a chromosome walking approach followed by Sanger sequencing, both of these sequence gap regions were closed to obtain a complete sequence coverage of *CEN5* as well. The sequences of these 11 centromere regions in the current GenBank

assembly were replaced by the newly assembled sequences as described above. The updated assembly was used for all of the analysis conducted in this study.

C. deuterogattii Oxford Nanopore sequencing and assembly

Samples for MinION sequencing were prepared as directed by the Oxford 1D genomic DNA sequencing protocol (v6). Briefly, 1.5 µg of high molecular weight genomic DNA was diluted into 46 µl of nuclease free water and pipetted into a g-TUBE (Covaris, Woburn). The sample was centrifuged for one min at 6000 rpm (3381 rcf). The g-TUBE was then inverted and centrifuged a second time. The sample was immediately removed by pipetting, and placed back into the original tube. The sheared DNA was combined with NEBNext FFPE RepairMix and buffer (NEB), mixed by inversion and incubated at 20ºC for 70 min. The reaction was cleaned by solid phase reversible immobilization (SPRI) using a 1X volume of AMPureXP magnetic beads (Beckman Coulter) and 70% ethanol for washes. The sample was eluted with nuclease free water and end repaired using the NEBNext Ultra II End-Repair/dA-tailing Module (NEB). The reaction was incubated for 5 min at 20ºC and then at 65ºC. The end prepped DNA was purified using SPRI as before, and then combined with the Oxford adapter mix (Oxford Nanopore Technologies kit SQK-LSK108) and NEB Blunt/TA Ligase Master Mix (NEB). Adapter ligation was performed for 10 min at room temperature and the reaction was cleaned using the modified SPRI detailed in the Oxford protocol. The sample was then immediately prepared for flow cell loading by combining with Oxford kit components RBF and LLB. 75 µl of the library mix was added dropwise to a primed MinION flow cell, and the sample was sequenced for 48 h. Base calling was performed using a full build of Albacore (version 2.0.2) after the sequencing run finished. Both raw and basecalled Nanopore reads are available in the NCBI SRA under accession SRP126415.

A total of 429,764 albacore-pass ONT fastq reads were assembled using Canu release v1.5 with the following parameters: -nanopore-raw <input.fastq>, correctedErrorRate=0.075, and stopOnReadQuality=false. After removing a small 1,797 artifactual contig of low complexity sequence, the assembly consisted of 15 contigs corresponding to the 14 chromosomes plus the mitochondria. The 15 contigs were polished by first aligning to them a total of 33,909,932 Illumina fragment paired reads using bwa mem (version 0.7.7-r441) followed by Pilon (version 1.13) correction using the --fix all setting. The polished contigs were aligned using nucmer (mummer package 3.23-64bit) to a *C. deuterogattii* (R265) PacBio assembly (27 contigs) and to a *Cryptococcus gattii* (WM276) Sanger assembly (14 chromosomes) to confirm chromosome structure. The mitochondrial contig was found to contain a duplicated region due to the circular configuration; an end to end overlapping region of 18684 bases was clipped from the 3' end resulting in a 31,190 base circular mitochondrial contig.

Contig ends were searched for telomeres using the known telomere motif sequence (TTAGGGG tandem repeats, allowing matches for TTAG [3,5]). For the 6 ends that were missing telomeric repeats, the contig end was extended by walking with aligned reads and then polishing. This was done by first aligning ONT reads (both raw and the canu-corrected set) to the polished contigs using bwa mem with parameter -x ont2d and identifying reads which aligned to a contig end and contained within the overhanging sequence the telomere motif. The consensus of aligned reads (between 1 and 3 reads identified matching each end) was added to extend the contig ends, followed by another round of Pilon correction using Illumina reads aligned to this updated assembly. In this updated assembly, an average of 45 bases (range of 13 to 97 bases) of telomeric repeat is present at each of the 28 scaffold ends. For the 5' end of scaffold 3.8, extended by one aligning ONT read, 3 copies of the telomeric repeat are located 81 bases from the end; the terminal sequence shares high similarity with

telomeric repeat arrays but contains more substitutions than is found at other ends, likely due in part to lower sequence quality from a single representative read.

ChIP-seq and bisulfite sequencing analysis

The ChIP-sequencing of *C. neoformans* mCherry-CENP-A as well as CENP-C-mCherry was done as previously described (4, 6). In total, 6 million single-end 36-nt reads (for CENP-C) or 10 million paired-end 100-nt reads (for CENP-A) were generated on the Illumina GAIIx platform. Raw reads were processed using SeqQC (version 2.2). The processed reads were aligned to the target *C. neoformans* genome using Geneious R9 software (http://www.geneious.com) (7). About 90% of the aligned reads were obtained per sample. All alignments for a particular read or pair were suppressed if more than 1000 reportable alignments existed for it. The alignments were further sorted into bam files. The graphs represented in Figure 1 and Figure S2 were generated using Integrative Genomics Viewer (IGV). It is notable that multiple breaks are observed in the binding patterns of both CENP-A and CENP-C (Figure 1B). These breaks could be due to technical limitations of the analysis where each read was allowed to align at multiple places. Due to this analysis criteria, a unique CENP-A bound region would appear as a dip if present in a repeat-rich region. However, this is unavoidable because these regions are highly repetitive in nature. We also tried to map single reads to unique regions, which led to poor read mapping and did not identify all 14 centromeres.

For *C. deuterogattii* CENP-C-mCherry ChIP-seq, ChIP data was generated using a HiSeq 2500 instrument to perform a 48 bp paired-end run. Reads were then aligned (using the same criteria applied for *C. neoformans*) to the *C. deuterogattii* genome using the short read component of the BWA aligner (8). The resulting alignment was converted, cleaned, and sorted using SAMtools (9) and Picardtools (https://broadinstitute.github.io/picard/). Peaks

were identified using the broad peaks setting of MACS2 (10). Bisulfite data of *C. neoformans*, acquired from a previously published study (PRJNA201680) (11), was aligned to the *C. neoformans* genome using Bismark v0.16.3 (12) in order to determine the proportion of methylation present at sites across the genome. All of the chromosome-wide read distribution and read depth graphs were generated using IGV (13, 14).

RNA extraction and real-time PCR assay

RNA was extracted from vegetatively growing cells of *C. neoformans* and *C. deuterogattii* as described previously (15). Briefly, the overnight culture was pelleted, washed with DEPC-treated water and resuspended in 1 ml of TRIzol reagent (Invitrogen, ThermoFisher Scientific). Glass beads (0.4 ml equivalent) were added into the tubes and vortexed, 4 cycles of 2 min each with 1 min interval. The RNA was then purified as per the TRIzol RNA extraction protocol provided by the manufacturer. The isolated RNA was subjected to DNase treatment, purified, and cDNA was prepared using oligodT primers (Sigma-Aldrich). Real-time PCR assays were performed using Tcn3, Tcn6 and the Clr4 gene (as control) specific primers (VYP183-188). The fold enrichment was calculated by double delta Ct method and was plotted using GraphPad Prism.

Methylation-specific PCR assay

Genomic DNA was isolated from overnight cultures *C. neoformans* and *C. deuterogattii* using the glass beads method described previously (16). The DNA was digested separately with CpG methylation-sensitive (HhaI, NciI and NotI) or insensitive (HindIII, PvuII and XhoI) enzymes for 14 h together with a no enzyme control reaction. The digested DNA was diluted 1:40 and used for PCR amplification. For PCR, two pairs of primers were designed for each *C. neoformans* (VYP75-76, VYP79-80) and *C. deuterogattii* (VYP741-

742, VYP743-744) - one pair amplifying centromere (*CEN*) DNA and another one for a non-centromeric (non-*CEN*) region. The PCR products obtained were visualized by gel electrophoresis using 0.8% agarose gels.

Transposon mapping analysis

The genomes of *C. neoformans* (H99), *C. deneoformans* (JEC21), and *C. deuterogattii* (R265) were scanned using the genome browser feature available in the FungiDB database (http://fungidb.org/fungidb/). The largest ORF-free regions with CENP-A or CENP-C binding on each chromosome were identified. For *Cryptococcus* species, the DNA sequence of each of the retrotransposons (Tcn1- Tcn6) has been previously reported (17). All of these sequences differ from each other with respect to their LTR regions while the domain architecture is conserved among them (17). The nucleotide sequences of these retroelements were harvested and used as query sequences in a BLASTn analysis (e value of 1) to identify all copies of transposable elements present in the genomes. The BLAST hits against each of the transposons in all chromosomes were obtained and mapped on each of the identified ORF-free regions. In case of overlapping mapping of different Tcn elements in the same region, the BLAST hit with longer sequence and lower e-value was considered while the other Tcn element hits were removed from the analysis.

Phylogenetic analysis of retrotransposon sequences was performed using MEGA6 (18). The full-length retrotransposon sequences were used for *C. neoformans*, *C. deneoformans* and *C. amylolentus*. For *C. deuterogattii*, the longest sequence traces of the Tcn elements were extracted and used for the analysis because full-length Tcn elements are missing from its genome. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (19). The tree with the highest log likelihood (-133683.3206) is shown in Figure S5A. Initial trees for the heuristic search were

obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach.

Genome synteny analysis

The genome comparative synteny analysis was performed using "SyMAP" using default parameters (http://www.agcol.arizona.edu/software/symap/) (20). The circular maps were generated using the circular map plugin available in the SyMAP software. Synteny analysis across the centromere regions among the three species was carried out using the synteny tool available in FungiDB (http://fungidb.org/) (21).

Experimental evolution

Experimental evolution was performed using *C. neoformans* wild-type (H99) and RNAi mutant derivatives (*rdp1*Δ and *ago1*Δ mutants). The strains were inoculated in 5 ml of YPD broth from a single colony and grown for 20 to 24 h at 30°C with shaking at 180 rpm. The next day, OD_{600} of the overnight culture was measured, and the required amount of cells were transferred into 5 ml of fresh YPD to achieve an initial OD_{600} of 0.1. This allowed enough inoculum of the culture for 24 h growth while not having an adverse effect on cells due to growth saturation or nutrient depletion. The culture was then further grown for 20-24 h, following which OD_{600} was again measured. The number of doublings was calculated for each of the strains from their initial (0.1) and final OD. On the next day, the overnight culture was again sub-cultured in fresh media starting with an initial OD_{600} of 0.1. Sub-culturing was continued on a daily basis until 1000 doublings were completed for each strain. DMSO stocks of each of the passaged cultures were made at regular intervals of 2 weeks, i.e. every 80-90 doublings.

Single colonies were streaked out from the 1000 doublings passaged strains of wild type, *ago1*Δ, and *rdp1*Δ mutants. Next, alterations of the centromere length were assessed by Pulsed-Field Gel Electrophoresis (PFGE) of genomic DNA. Plugs were prepared from single colonies as previously described (22), and digested overnight with the restriction enzyme NotI-HF (NEB) and then run in 1% agarose gel in 0.5X TBE with a switching time 7 – 60 s, for 120 h at 14°C using a CHEF apparatus. The enzyme was chosen such that the entire centromere region is released as a single fragment along with flanking sequences that can be used as a probe. The DNA was then transferred to a membrane, and hybridized with probes targeting chromosomal regions flanking the centromeres, as previously described (22). One colony for each *rdp1*Δ-1000 and *ago1*Δ-1000 strain that showed changes in *CEN2* compared to wild-type-1000 was used for PacBio sequencing. Single colonies were also streaked out from the 0 doubling strains (wild-type-0, *rdp1*Δ-0 and *ago1*Δ-0) and one colony from each was used for PacBio sequencing. The genomes were assembled *de novo* using Canu and each centromere length was measured as the intergenic region between the centromere flanking ORFs. The transposon mapping in the new assemblies was done by BLASTn analysis using Tcn1-Tcn6 DNA sequences.

Prediction of centromeres in Ustilago species

A previous study in *U. maydis* predicted its centromeres based on the presence of a transposon (HobS)-rich sequence as well as plasmid stability assays (23). We performed RNA-seq analysis for *U. maydis* using the transcriptome data available from a previous study (24). For RNA-seq analysis, the reads were aligned to the *U. maydis* reference genome using Geneious R9 software and plots for each chromosome were generated. Combining the earlier prediction with the lack of polyA RNA reads, one region on each chromosome was identified as the putative centromere. Synteny with the *U. maydis* genome and RNA-seq (25) analysis

were performed using Geneious R9 software to predict putative centromeres in *U. bromivora* as well.

For *U. hordei*, neither RNA-seq nor synteny analysis could be performed due to the lack of a suitable chromosome-wide assembly as well as RNA-seq data. Thus, as an alternative approach, an *U. hordei* BAC clone library was utilized to measure the length of putative centromeres (26, 27). First, a BLAST analysis with end sequences of BAC clones against *U. maydis* genome was performed. Considering that centromere flanking regions between all three *Ustilago* species are syntenic, we identified the BAC clones that harbor the putative *U. hordei* centromeres. Based on the size of the BAC clones, the length of the cloned syntenic region in every BAC clone was estimated. The length of the sytenic region (based on BLAST hits) in *U. maydis* genome was also measured, and the difference between the length of syntenic regions from *U. maydis* and *U. hordei* was calculated. Because the genomic content between two species is similar, the difference in length was attributed to increased centromere length in *U. hordei*. By this approach, the length of 17 centromeres out of 23 in *U. hordei* was predicted. Next, PacBio sequencing was performed for *U. hordei* followed by *de novo* assembly of the *U. hordei* genome. Synteny analysis was performed using the refined genome, and 18 putative centromeric regions were identified. Fifteen of these identified putative centromeres were the same as the ones predicted using the BAC based approach and showed a consensus on centromere length. Two regions identified using the BAC clone approach are broken in our current PacBio assembly whereas three regions identified using the PacBio approach lack equivalent BAC clones. Thus combining the data from BAC clone inserts and the PacBio assembly, we could determine the length of 20 centromeres in *U. hordei*.

Supplementary figure legends

Figure S1. Identification of CENP-A and CENP-C and subcellular localization of CENP-A in the *Cryptococcus* species complex. (A) Alignment of CENP-A proteins from *C. deuterogattii* (Cdg, ORF no. CNBG_0491), *C. neoformans* (Cn, CNAG_00063), and *C. deneoformans* (Cdn, CNA00540) with CENP-A sequences of *Drosophila melanogaster* (Dm), *Mus musculus* (Mm), *Homo sapiens* (Hs), *Saccharomyces cerevisiae* (Sc), *Ustilago maydis* (Um), *Ustilago bromivora* (Ub), *Ustilago hordei* (Uh), *Candida albicans* (Ca), *Neurospora crassa* (Nc) and *Schizosaccharomyces pombe* (Sp). The C-terminal region of CENP-A carries the conserved histone-fold domain (HFD). **(B)** Multiple sequence alignment of CENP-C proteins in *C. deuterogattii* (CNBG_4461), *C. neoformans* (CNAG_05391) and *C. deneoformans* (CNH00580) with other species revealed conservation of the CENP-C box and the DNA binding "Cupin" domain in these three species. **(C)** The sub-cellular localization patterns of a conserved kinetochore protein CENP-A at various cell cycle stages (interphase, pro-metaphase, and anaphase) in *C. neoformans*, *C. deneoformans*, and *C. deuterogattii*. Bar, 5 μm.

Figure S2. CENP-A and CENP-C bound centromeres are associated with specifically modified histone H3 and DNA in *C. neoformans*. ChIP-seq analysis identified centromeres as overlapping binding sites of CENP-A and CENP-C identified the location of the centromeres on each chromosome of *C. neoformans*. H3K9diMe, H3K27diMe ChIP-seq data (28) and bisulfite sequencing data (11) were reanalyzed to determine sites of respective histone marks and DNA methylation across the chromosomes in *C. neoformans* and found to be enriched at the centromeres on each chromosome. The retrotransposons (Tcn1-Tcn6) were also mapped along the length of the chromosomes and found to be enriched at the centromeres. The additional CENP-A peak, appearing on chromosome 2, is probably an experimental artifact because the peak is also present at the same region in "Input DNA" control as well and hence it was considered to be a false positive peak. The extra peak in chromosome 11 lies in a gap region that shows similarity to centromeric retroelements and hence may be due to a genome assembly error.

Figure S3. Centromeres in *Cryptococcus* species complex are poorly transcribed. RNA-seq reads were obtained from NCBI (SRR3199613 for *C. neoformans*, SRR1796479 for *C. deneoformans*, SRR5209627 for *C. deuterogattii*), aligned to respective reference genomes

for all three of the *Cryptococcus* species and centromere regions were probed for the presence/absence of RNA transcripts. The regions shown here include centromeres together with 50 kb centromere flanking chromosomal regions on both sides. The centromere regions are highlighted in red bars while the black peaks mark the RNA-seq reads.

Figure S4. Identification of centromeres in *C. deneoformans* and *C. deuterogattii*. (A) A circular map showing synteny between the *C. neoformans* and *C. deneoformans* genomes. **(B)** A chromosome-wide map showing the location of centromeres along with the distribution of retrotransposon elements in *C. deneoformans*. The regions amplified for ChIP-qPCR analysis are also marked in the maps. **(C)** Synteny analysis between the *C. neoformans* and *C. deuterogattii* genomes showed a number of chromosomal rearrangements between the two species (See Supplementary table S2 for details). **(D)** CENP-C (mCherry)-ChIP-seq analysis identified locations of centromeres in *C. deuterogattii* genome. The retrotransposon (Tcn1-Tcn6) locations along the chromosomes are also shown.

Figure S5. Phylogenetic analysis of retrotransposons in the *Cryptococcus* species complex. (A) A phylogenetic tree drawn to show the evolutionary relationship between retrotransposons present in *C. neoformans* (Tcn1.H99-Tcn6.H99), *C. deneoformans* (Tcn1.JEC21-Tcn6.JEC21), *C. deuterogattii* (Tcn1.R265-Tcn6.R265), and *C. amylolentus* (Tcen1-Tcen6). The tree is drawn to scale with branch lengths measured in the number of substitutions per site. **(B)** A graph showing the expression levels of two Tcn elements, Tcn3 and Tcn6, in *C. neoformans* and *C. deuterogattii* as compared to a control gene region, Clr4.

Figure S6. DNA methylation at the centromere is lost in *C. deuterogattii*. (A) The *DNMT5* ORF is truncated in *C. deuterogattii* at the syntenic locus to that of *C. neoformans* and *C. deneoformans*. **(B)** The diagram to show the rationale of the assay used to determine DNA methylation status at the centromere. **(C)** PCR analysis revealed a lack of methylation at the centromere DNA in *C. deuterogattii* unlike that of *C. neoformans*. '+' or '-' refers to the presence or absence of the restriction site of a specific enzyme respectively. Enzymes in red letters are CpG methylation-sensitive while others are not. One centromeric region (*CEN6* for *C. neoformans* and *CEN9* for *C. deuterogattii*) and one non-centromeric region (Chr1: 1726512-1727921 for *C. neoformans* and SC6: 376524-377415 for *C. deuterogattii*) was subjected to the assay. **(D)** PacBio sequencing based base-modification analysis revealed a

high level of DNA methylation at centromeres in *C. neoformans* but a much reduced level in *C. deuterogattii*. The grey shaded regions represent the centromere of each chromosome.

Figure S7. Centromeres in the *Ustilago* species complex. Centromeres were predicted in *U. maydis* based on features including the presence of transposons, lack of transcription, and a long stretch of an ORF-free region. All of the predicted centromeres are rich in HobS retroelements but poorly transcribed as revealed by the absence of polyA-RNA. Centromeres were identified in *U. bromivora* and *U. hordei* by synteny analysis with the *U. maydis* genome. The putative centromeres in these two species are also poorly transcribed as shown by the lack of polyA-RNA from these regions. RNA-seq reads were obtained from EBI or NCBI (ERR184024 for *U. maydis*, SRR4381675 for *U. bromivora*, SRR5235715 for *U. hordei*).

Figure S8. RNAi provides stability of retrotransposon-rich genomic loci. (A) Schematic of experimental evolution performed using *C. neoformans* wild-type *(RDP1 AGO1*), *rdp1*Δ, and *ago1*Δ mutants. P in P_1, P_2, P_n refers to passage numbers that were made on a daily basis (see Supplementary Material and Methods for details). **(B)** Southern blot analysis of PFGE gel revealed genomic rearrangements at the centromeres when RNAi mutant (*rdp1*Δ and *ago1*Δ) strains were passaged for 1000 generations. Red stars indicate the length of NotI fragments expected in the wild-type strain *C. neoformans* that was also passaged for 1000 generations. EtBr refers to ethidium bromide stained gels while *CEN2* refers to blots developed using a probe against the *CEN2* region. The probe location is shown with respect to the centromere location in the map below. **(C)** PacBio sequencing followed by synteny analysis of centromeric regions revealed genomic rearrangements at the centromeres in RNAi mutant (*rdp1*Δ and *ago1*Δ) strains as compared to the wild-type strain, all passaged for 1000 doublings. While the change in *CEN7* was observed only after 1000 doublings, the *CEN2* length was reduced in unpassaged RNAi mutant strains as well. See Supplementary table S5 for more details.

Supplementary table S1. The centromere coordinates in *C. neoformans*, *C. deneoformans,* and *C. deuterogattii*.

CEN #	*C. neoformans*	*C. deneoformans*	*C. deuterogattii**
1	Chr1: 970169-1006931 (36763)	NC_06670: 937505-998182 (60678)	Chr1: 1119536-1140609 (21074)
2	Chr2: 835384-889427 (54044)	NC_06684: 855280-905374 (50095)	Chr2: 686138-699669 (13532)
3	Chr3: 1370568-1409632 (39065)#	NC_06680: 139615-178627 (39013)	Chr3: 1246085-1264856 (18772)
4	Chr4: 708804-752337 (43534)	NC_06681: 129330-176311 (46982)	Chr4: 912396-924705 (12310)
5	Chr5: 1559983-1587231 (27248)	NC_06686: 220960-273717 (52758)	Chr5: 494400-512261 (17862)
6	Chr6: 780649-821756 (41108)	NC_06687: 777728-854403 (76676)	Chr6: 577407-592594 (15188)
7	Chr7: 525714-584338 (58625)	NC_06691: 863695-936334 (72640)	Chr7: 412760-421530 (8771)
8	Chr8: 451162-512653 (61492)	NC_06692: 882181-912116 (29936)	Chr8: 816771-828467 (11697)
9	Chr9: 801830-839446 (37617)	NC_06694: 323826-388577 (64752)	Chr9: 753677-767446 (13770)
10	Chr10: 199434-243741 (44308)	NC_06679: 802162-882405 (80244)	Chr10: 361045-370399 (9355)
11	Chr11: 868824-933658 (64835)#	NC_06685: 801507-911882 (110376)	Chr11: 555194-569391 (14198)
12	Chr12: 139633-171048 (31416)	NC_06682: 122048-182012 (59965)	Chr12: 557411-571509 (14099)
13	Chr13: 579772-632362 (52591)	NC_06683: 569940-644450 (74511)	Chr13: 105564-120756 (15193)
14	Chr14: 441845-477986 (36141)#	NC_06693: 706065-761098 (55034)	Chr14: 196268-217925 (21658)

The numbers in brackets denote the length of the centromere in basepair (bp).
*The chromosome number are noted as per our latest chromosome-wide assembly of *C. deuterogattii*.
Centromeres with gaps and hence the actual length may be longer than estimated here.

Supplementary table S2. Chromosome level synteny between *C. neoformans*, *C. deneoformans*, and *C. deuterogattii*.

C. neoformans	*C. deneoformans*	*C. deuterogattii*
Chr1	NC_006670	Chr3 + Chr4
Chr2	NC_006684	Chr3 + Chr4
Chr3	NC_006685 + NC_006680	Chr1 + Chr11
Chr4	NC_006681 + NC_006693	Chr2 + Chr9 + Chr14
Chr5	NC_006686	Chr2 + Chr9 + Chr14
Chr6	NC_006687	Chr6
Chr7	NC_006691	Chr5
Chr8	NC_006692	Chr7
Chr9	NC_006694	Chr8
Chr10	NC_006679	Chr2 + Chr9 + Chr14
Chr11	NC_006685 + NC_006680	Chr1 + Chr11
Chr12	NC_006682	Chr12
Chr13	NC_006683	Chr13
Chr14	NC_006693	Chr10

Supplementary table S3. The centromere coordinates in *U. maydis*, *U. bromivora*, and *U. hordei*.

CEN#	***U. maydis* coordinates**	***U. bromivora* coordinates**	***U. hordei* coordinates***
1	Chr1: 672652-681079 (8428)	Chr2: 452858-488889 (36032)	Sc2: 1461689-1494602 (32914)
2	Chr2: 1723075-1739353 (16279)	Chr3: 1503289-1530560 (27272)	Sc3: 132303-174333 (42031)
3	Chr3: 446528-483667 (37140)	Chr4: 1374332-1396569 (22238)	N.D
4	Chr4: 67583-79380 (11798)	Chr5: 1889935-1915550 (25616)	Sc12: 396214-427879 (31666)
5	Chr5: 627251-639414 (12164)	Chr1: 480046-509128 (29083)	N.D.
6	Chr6: 921977-929042 (7066)	Chr6: 977201-998292 (21092)	Sc4: 1238539-1273654 (35116)
7	Chr7: 838836-845933 (7098)	Chr7: 76047-109059 (33013)	Sc6: 993198-1027945 (34748)
8	Chr8: 171427-191010 (19584)	Chr8: 705381-726397 (21017)	Sc14: 188704-229586 (40883)
9	Chr9: 142767-149931 (7165)	Chr10: 115367-133510 (18144)	Sc16: 576749-612953 (36205)
10	Chr10: 131074-139218 (8145)	Chr14: 100466-137856 (37391)	Sc12: 1237939-1276412 (38474)
11	Chr11: 258406-267040 (8635)	Chr15: 418425-445473 (27049)	Sc11: 1148546-1180655 (32110)
12	Chr12: 73829-95366 (21538)	Chr12: 26281-53109 (26829)	N.D.
13	Chr13: 368081-397514 (29434)	Chr11: 263218-289903 (26686)	N.D.
14	Chr14: 357506-373488 (15983)	Chr13: 352996-384709 (31714)	Sc10: 306600-340267 (33668)
15	Chr15: 262492-270546 (8055)	Chr19: 238674-268065 (29392)	Sc11: 259150-294933 (35784)
16	Chr16: 410262-420966 (10705)	Chr17: 109402-139667 (30266)	Sc29: 197725-223379 (25655)
17	Chr17: 90877-106428 (15552)	Chr16: 63155-78285 (15131)	Sc7: 138981-170485 (31505)
18	Chr18: 70385-89661 (19277)	Chr20: 55817-83227 (27411)	Sc15: 76204-115173 (38970)
19	Chr19: 526864-545420 (18557)	Chr18: 506603-548848 (42246)	Sc18: 503310-570065 (66756)
20	Chr20: 475200-479322 (4123)	Chr9: 707407-732237 (24831)	N.D.
21	Chr21: 321332-331632 (10301)	Chr21: 317190-350653 (33464)	Sc20: 385526-421512 (35987)
22	Chr22: 169493-183956 (14464)	Chr22: 160000-185505 (25506)	Sc22: 300552-345151 (44600)
23	Chr23: 232483-260749 (28267)	Chr23: 31187-58282 (27096)	Sc24: 324255-362781 (38527)

Um and Ub chromosome numbers are as per available in NCBI genome assemblies. Uh coordinates are as per our Pac-Bio assembly. The numbers in brackets denote the length of the centromeres in basepair.

*in the absence of a chromosome-wide genome assembly of *U. hordei*, scaffold (Sc) numbers are noted; N.D., Not determined.

Supplementary table S4. Status of RNAi genes and DNA methylation in various fungi.

Species	Ago	Dcr	Rdp	DNAme
Neurospora crassa	✓	✓	✓	✓
Fusarium graminearum	✓	✓	✓	✓
Ustilago hordei	✓	✓	✓	✓
Ustilago bromivora	✓	✓	✓	✓
Cryptococcus amylolentus	✓	✓	✓	N.D.
Ustilago maydis	×	×	×	×
Cryptococcus deneoformans	✓	✓	✓	✓
Cryptococcus neoformans	✓	✓	✓	✓
Cryptococcus deuterogattii	×	✓	×	×
Candida tropicalis	✓	✓	×	N.D.
Candida albicans	✓	✓	×	×
Schizosaccharomyces pombe	✓	✓	✓	×
Clavispora lusitaniae	✓	✓	×	N.D.
Schizosaccharomyces octosporus	✓	✓	✓	×
Schizosaccharomyces japonicus	✓	✓	✓	×
Komagataella phaffii	×	×	×	N.D.

✓, present; ×, absent; N.D., Not determined

Species shaded in grey represent the species that have lost one or more RNAi machinery gene.

Supplementary table S5. The generation time for the strains used in this study.

Strain name	Generation time (min ± standard deviation)
C. neoformans (wild-type) – 0 doubling	94.6 ± 2.4
YPH351 (*rdp1*Δ) – 0 doubling	91.6 ± 5.7
YSB299 (*ago1*Δ) – 0 doubling	90.0 ± 3.2
C. neoformans (wild-type) – 1000 doublings	86.7 ± 0.8
YPH351 (*rdp1*Δ) – 1000 doublings	90.0 ± 0.2
YSB299 (*ago1*Δ) – 1000 doublings	87.7 ± 0.8
C. neoformans	96.3 ± 3.8
C. deneoformans	97.5 ± 3.2
C. deuterogattii	83.6 ± 1.2

Supplementary Table S6. Centromere lengths in experimentally evolved strains.

	C. neoformans wild-type – 0 doublings	*C. neoformans* wild-type – 1000 doublings	*ago1*Δ – 0 doublings	*ago1*Δ – 1000 doublings	*rdp1*Δ - 0 doubling	*rdp1*Δ – 1000 doubling
CEN1	36540	36769	36770	36768	36770	36769
CEN2	60735	60735	**54046**	**54048**	**54049**	**54048**
CEN4	43560	43561	43561	43558	43563	43557
CEN5	41030	41031	41030	41031	41032	41031
CEN6	41114	41114	41112	41040	41115	41114
CEN7	53942	53940	53943	53943	53943	**44519**
CEN9	37624	37624	37625	37624	37625	37624
CEN11	64189	64187	64186	64193	64190	64193
CEN12	31423	31437	31429	31432	31428	31425
CEN13	52611	52610	52603	52608	52612	52609
CEN14	52484	52494	52495	52493	52495	52495

The numbers denote the centromere lengths in base pairs (bp).
CEN3, *CEN8* and *CEN10* are not listed because they were not covered completely in one or more strains analyzed in the experiment.

Supplementary table S7. Strains used in this study.

Strain	Genotype	Source
C. neoformans		
H99	α wild type	(29)
KN99	**a** wild type	(30)
CNVY101	**a** *mCherry-CSE4::NEO*	(1)
CNVY102	**a** *MIF2-mCherry::NEO*	(1)
YPH351	α *rdp1*Δ*::NEO*	(15)
YSB299	α *ago1*Δ*::NAT*	(15)
CNVY251	H99 – 1st doubling	This study
CNVY253	YSB299 – 1st doubling	This study
CNVY256	YPH351 – 1st doubling	This study
CNVY263	H99 – 1000 doublings	This study
CNVY266	YSB299 – 1000 doublings	This study
CNVY268	YPH351 – 1000 doublings	This study
SS-E629	H99 – 1000 doublings – colony 1	This study
SS-E640	YSB299 – 1000 doublings – colony 6	This study
SS-E643	YPH351 – 1000 doublings – colony 3	This study
C. deneoformans		
JEC21	α *wild type*	(31)
CNVY501	α *mCherry-CSE4::HYG*	This study
CNVY502	α *MIF2-mCherry::NEO*	This study
C. deuterogattii		
R265	α *wild type*	(32)
CNVY701	α *mCherry-CSE4::NEO*	This study
CNVY702	α *MIF2-mCherry::NEO*	This study

Supplementary table S8. Primers used in this study.

Name	Sequence (5' --- 3')	Purpose
VYP75	AGTCTCGTGTGGCTATGATT	*C. neoformans CEN* methylation
VYP76	GGATCTGCTTGACAGTGTCA	
VYP79	CCAACCGAAGCCCAAGACAA	*C. neoformans* non-*CEN* methylation
VYP80	TTGAAGGATGATCCGGCCGA	
VYP501	GGATAGAGCAAGATCTGCTAGGTC	*C. deneoformans* CENP-C-mCherry tagging
VYP502	CTCGCCCTTGCTCACCATTCTCCTGCTCTTCCCCTTAC	
VYP503	GTAAGGGGAAGAGCAGGAGAATGGTGAGCAAGGGCGAG	
VYP504	GCTTCGTTACTGACAACAATATATCCCAAGCTTGGTACCGAGCTC	
VYP505	GAGCTCGGTACCAAGCTTGGGATATATTGTTGTCAGTAACGAAGC	
VYP506	TGAGGCAGGAATCATGTAGTC	
VYP507	GGCTGCGCTGTTATCAAGGAGATC	
VYP508	CTTGGGAGGGACGAATACATTGACCTG	
VYP509	GTGCAACTGCTATGTAGCTG	*C. deneoformans CEN1* ChIP qPCR – 1
VYP510	TGTGGAACGTCTGACAGTG	
VYP511	CTTATGCTCCTTCAAGTGC	*C. deneoformans CEN1* ChIP qPCR – 2
VYP512	ACCCAGCCTTGCTACTCAC	
VYP513	CTACCTTCTTCGACATTGGC	*C. deneoformans CEN2* ChIP qPCR – 1
VYP514	CCATCAAGTCGCCAAGTGC	
VYP515	ATCGGCAAGCACTAGTAGC	*C. deneoformans CEN2* ChIP qPCR – 2
VYP516	ACGTCATGACAGACCATGC	
VYP517	GTGGTCAATACGCAAGTCG	*C. deneoformans CEN3* ChIP qPCR – 1
VYP518	ACCGACCACTTCACTCTC	
VYP519	CAGTAGACTGATCAGCAAGC	*C. deneoformans CEN3* ChIP qPCR – 2
VYP520	GCCACAATGACATACGAGC	
VYP521	CGTCTTCGCTATTCCAGTTC	*C. deneoformans CEN4* ChIP qPCR – 1
VYP522	CGTGACATTGTTCAGAGC	
VYP523	CAACAAGGGGAATAGGAAGG	*C. deneoformans CEN4* ChIP qPCR – 2
VYP524	GCTGATCGATGGACTCTTG	
VYP571	TCGTCGAGCCGCATATGC	*C. deneoformans CEN5* ChIP qPCR – 1
VYP526	ACACTCCAGCGAAAATTGC	
VYP572	GTGTTGCTTGCGTCGGTG	

VYP528	TGAAGGAAATGGTGGCACG	*C. deneoformans CEN5* ChIP qPCR – 2
VYP529	ACCAGCACCAGTCGCTTC	*C. deneoformans CEN6* ChIP qPCR – 1
VYP530	GTCTCAGACTTCATTCTCATC	
VYP531	CATAACTCGACTTCAACTCG	*C. deneoformans CEN6* ChIP qPCR – 2
VYP532	CCTTGACATCCGCACCAG	
VYP533	AACATCTTGGTGACTGTCG	*C. deneoformans CEN7* ChIP qPCR – 1
VYP534	AAACCATCTATCTTGAAGCAC	
VYP535	AGCACGGAAATCGCAGAC	*C. deneoformans CEN7* ChIP qPCR – 2
VYP536	TGAATGCAGGACGTCTTCG	
VYP576	CATTCTCACCATATGGTAGG	*C. deneoformans CEN8* ChIP qPCR – 1
VYP577	CGCGTTTCGGTGAAGTCC	
VYP578	TTGGGTGCAGTGGTTTGTGC	*C. deneoformans CEN8* ChIP qPCR – 2
VYP579	CAAGGCAGGGAAGGTAGC	
VYP541	AATTGATAGGAACACTGATCAG	*C. deneoformans CEN9* ChIP qPCR – 1
VYP542	TACAGTCACAAGTACCTTGC	
VYP543	ACAACGCAGTAGTTCAAGTG	*C. deneoformans CEN9* ChIP qPCR – 2
VYP544	CCCCGAAGTACTAACCTTGC	
VYP545	TCAGACCCATCGTCAATCATG	*C. deneoformans CEN10* ChIP qPCR – 1
VYP546	CGAAGCCGATGCTGAGTAC	
VYP547	TCGGTTGAATTCCCTCCTG	*C. deneoformans CEN10* ChIP qPCR – 2
VYP548	ATGACTGTCTTGTTAGATCG	
VYP569	TCACTGGATTCTTTGACAAGG	*C. deneoformans CEN11* ChIP qPCR – 1
VYP550	CTGCTCTTGGATGATATAGGAC	
VYP551	GTCTAGAGAGAGCTTGAGC	*C. deneoformans CEN11* ChIP qPCR – 2
VYP570	CCAGAACACTTACAATATCGAAAC	
VYP553	GTCGAGTAGGCGAGGAAC	*C. deneoformans CEN12* ChIP qPCR – 1
VYP554	ACCTCAACACAGTCGACG	
VYP555	TAGGCGGTGTTGACGACAG	*C. deneoformans CEN12* ChIP qPCR – 2
VYP556	TCATTGGTGACACTACCTAC	

VYP557	AGTCACACGTCATACAAGTC	*C. deneoformans CEN13* ChIP qPCR – 1
VYP558	AACCTAGGAACTCTACTGAG	
VYP559	ACGACAATCGTAGCATCG	*C. deneoformans CEN13* ChIP qPCR – 2
VYP560	CTATGTCCTACTATCACGAC	
VYP561	CGTTCGTGGTATAGGTCTAGAG	*C. deneoformans CEN14* ChIP qPCR – 1
VYP567	CCATTGCTAGTTTCGCATC	
VYP563	CATCCTTCCCCCATATGATG	*C. deneoformans CEN14* ChIP qPCR – 2
VYP568	TCAACAGCGTCGCATTAATG	
VYP573	CTACTCATACAACGACACCTC	*C. deneoformans* non-*CEN* ChIP qPCR
VYP566	TGAGTGAGCCACCTATAACG	
VYP701	ACGTCCGTCCGAACTTGG	*C. deuterogattii* CENP-C-mCherry tagging
VYP702	CTCGCCCTTGCTCACCATTCTCCTACTCTTCCCTTTACTTTTTCTC	
VYP703	GAGAAAAAGTAAAGGGAAGAGTAGGAGAATGGTGAGCAAGGGCGAG	
VYP704	CATCTTCCCCCTGCCAGTCCAAGCTTGGTACCGAGCTC	
VYP705	GAGCTCGGTACCAAGCTTGGACTGGCAGGGGGAAGATG	
VYP706	AGCCGCCAGGTAGATGAGG	
VYP707	CACTATCCCTGAAGATCCACCTATACC	
VYP708	CGATTGCCTGTTTCACTTCACTC	
VYP741	CTGACCTCTAGTTGCAGGAGC	*C. deuterogattii CEN* methylation
VYP742	CCAGATGATGTGGCATTCAAG	
VYP743	CTTGTCTCGGCGTCCCAAAC	*C. deuterogattii* non-*CEN* methylation
VYP744	AAAACGCTCAAAGCCTCTACG	
VYP183	GACAGGGTGGACTTGGTC	qPCR primers for Tcn3 expression
VYP184	GATGCTGTCAAGGCAGG	
VYP185	TACCAGCTAGCTTCTGG	qPCR primers for Tcn6 expression
VYP186	GCTGGTATGGCAAGAA	
VYP187	GTCAAGGACTTTCATCC	qPCR primers for Clr4 expression
VYP188	ATACCCTTGTAACTGATAC	
JOHE41845/SS	CGCAGAAAAGAGACATCGGC	*C. neoformans CEN2* Southern probe
JOHE41846/SS	GGCTTGCAAATGCACTGGGT	

1

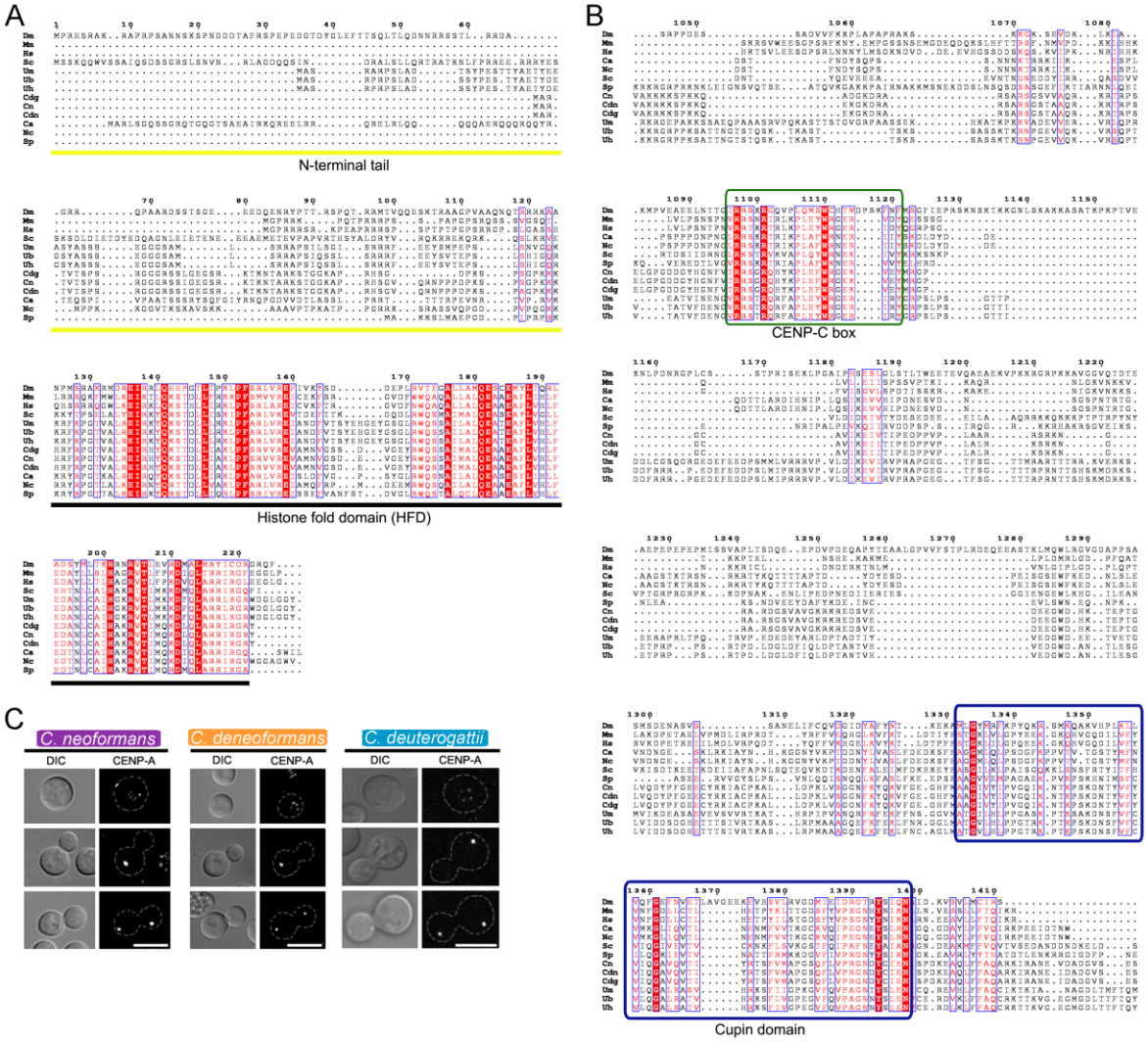
2

Supplementary table S9. NCBI data submissions related to this study.

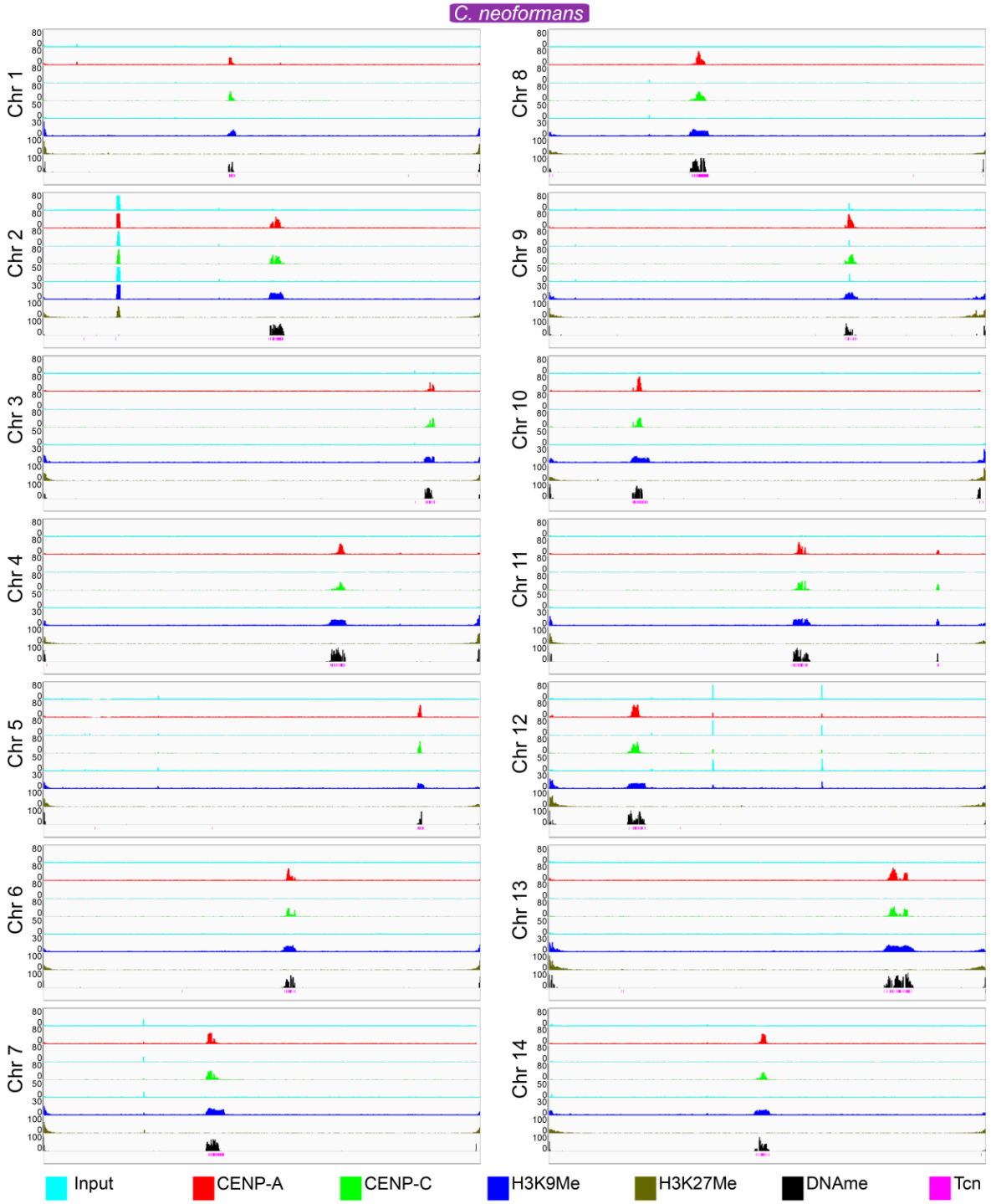
Sr. no.	Experiment	Sample	Reads obtained	Reads aligned (%)	Accession number
1.	*C. neoformans* CENP-A ChIP-seq	Input DNA	2,76,08,910	2,60,03,094 (94.2)	SRS2381716
2.	*C. neoformans* CENP-A ChIP-seq	IP DNA	8,27,94,252	7,20,05,466 (86.9)	SRS2381718
3.	*C. neoformans* CENP-C ChIP-seq	Input DNA	5,33,98,921	4,88,01,578 (91.4)	SRS2381717
4.	*C. neoformans* CENP-C ChIP-seq	IP DNA	1,20,09,376	1,12,65,782 (93.8)	SRS2381719
5.	*C. deuterogattii* CENP-C ChIP-seq	Input DNA	3,93,55,378	3,83,37,884 (97.4)	SRS2381717
6.	*C. deuterogattii* CENP-C ChIP-seq	IP DNA	2,55,48,638	66,84,103 (26.1)	SRS2381719
7.	*C. neoformans* PacBio sequencing	genomic DNA	3,50,846	N.A.	SRS2403243
8.	*C. deuterogattii* PacBio sequencing	genomic DNA	1,25,502	N.A.	SRS2403242
9.	*U. hordei* PacBio sequencing	genomic DNA	2,45,303	N.A.	SRS2403241
10.	*C. deuterogattii* Nanopore sequencing	genomic DNA	4,29,764	N.A.	SRS2747819
11.	*C. deuterogattii* chromosome-wide assembly	N.A	N.A	N.A	SAMN08330675 (R265_Chr00000000)
12.	*C. neoformans* PacBio assembly	N.A.	N.A.	N.A.	NPNB00000000
13.	*C. deuterogattii* PacBio assembly	N.A.	N.A.	N.A.	NPNA00000000
14.	*U. hordei* PacBio assembly	N.A.	N.A.	N.A.	NPMZ00000000
15.	*C. neoformans*-WT-0 doubling PacBio sequencing	genomic DNA	2,17,583	N.A.	SRS2803057
16.	*C. neoformans* -ago1null-0 doubling PacBio sequencing	genomic DNA	3,71,570	N.A.	SRS2803054
17.	*C. neoformans* -rdp1null-0 doubling PacBio sequencing	genomic DNA	6,36,112	N.A.	SRS2803056
18.	*C. neoformans* -WT-1000 doubling PacBio sequencing	genomic DNA	7,48,909	N.A.	SRS2803055
19.	*C. neoformans* -ago1null-1000 doubling PacBio sequencing	genomic DNA	7,52,306	N.A.	SRS2803052
20.	*C. neoformans* -rdp1null-1000 doubling PacBio sequencing	genomic DNA	8,12,833	N.A.	SRS2803053

N.A., Not applicable

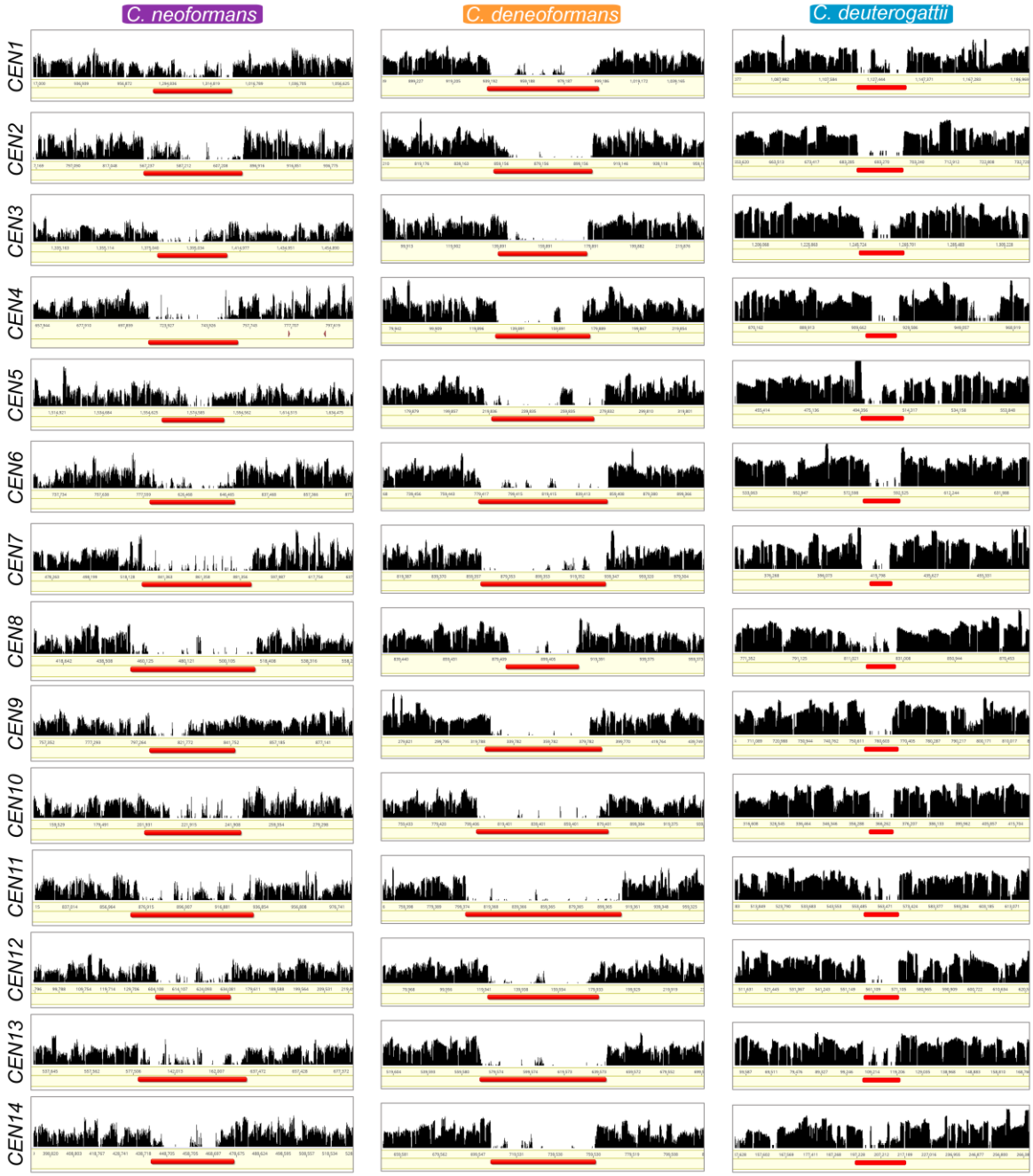
Supplementary figure S1



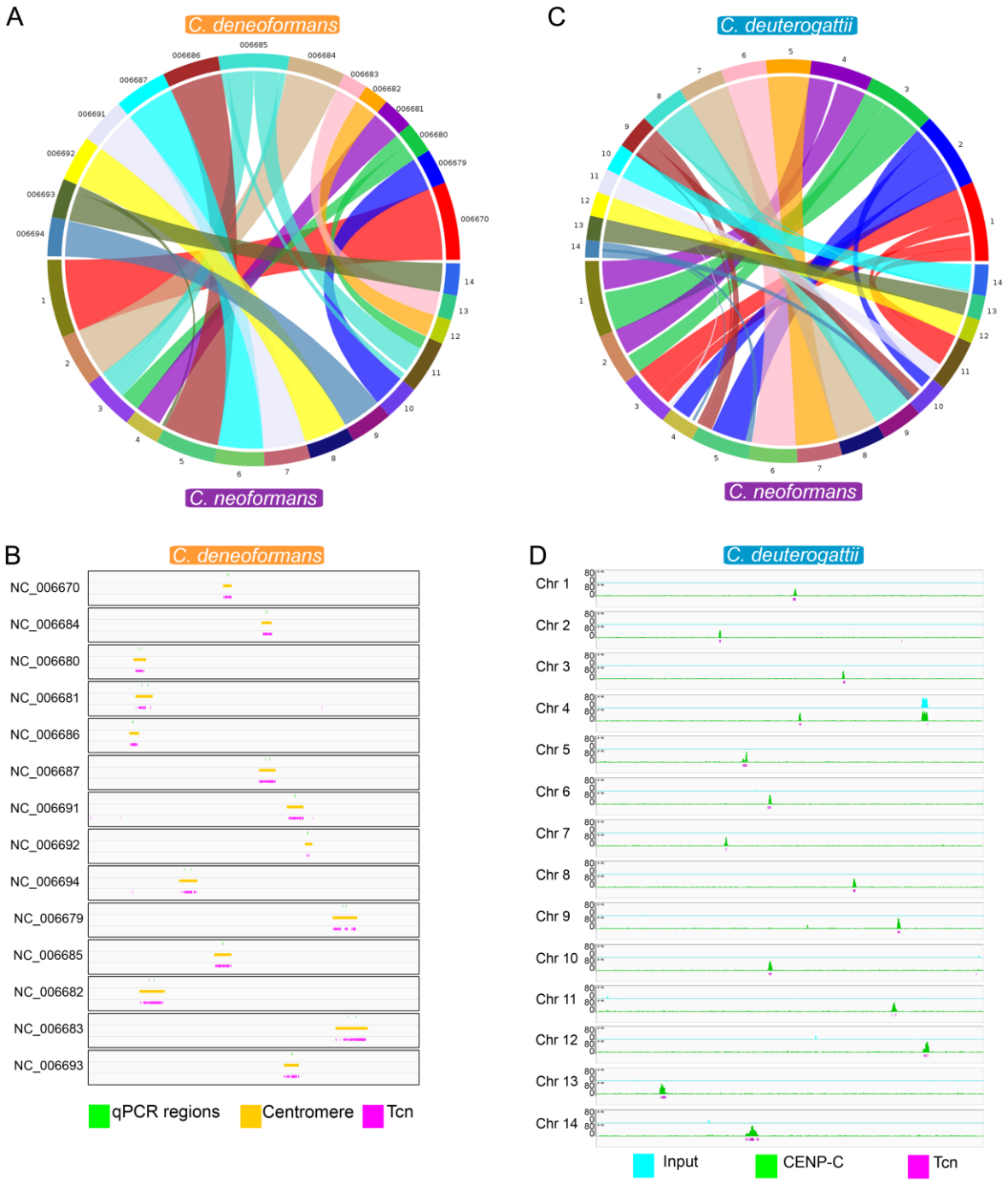
Supplementary figure S2



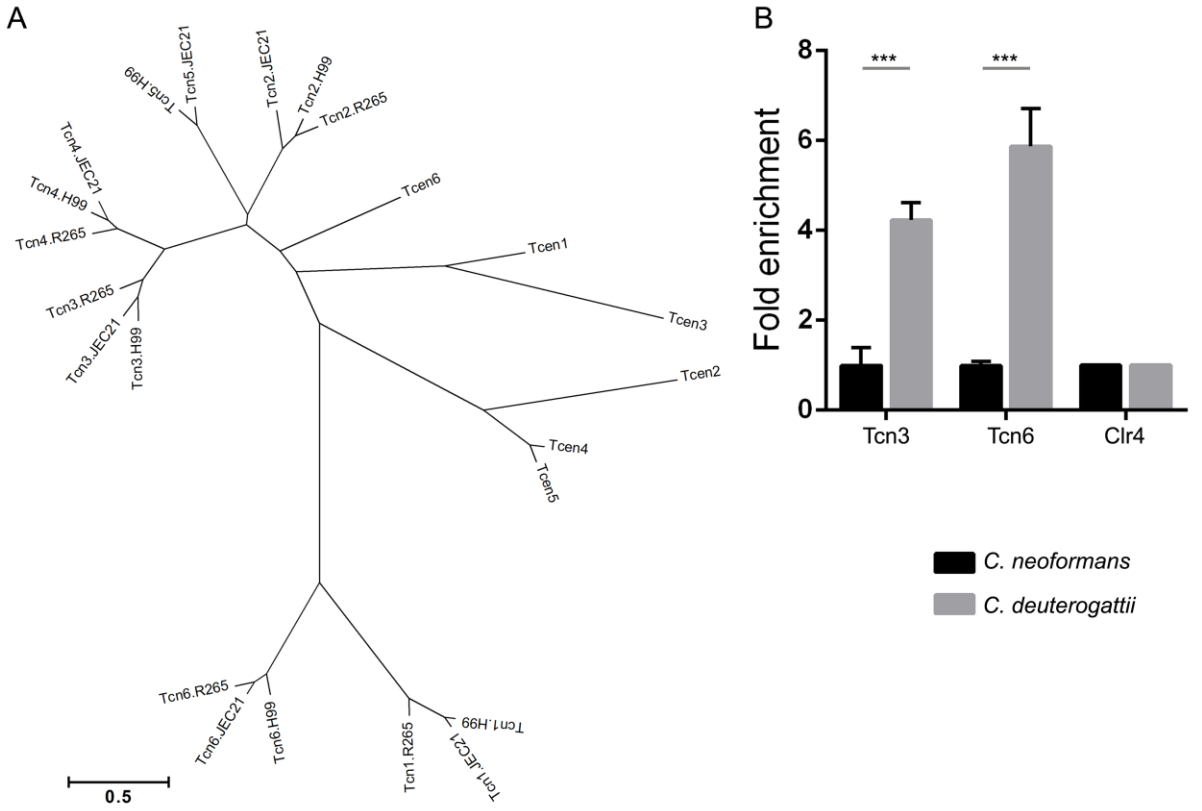
Supplementary figure S3



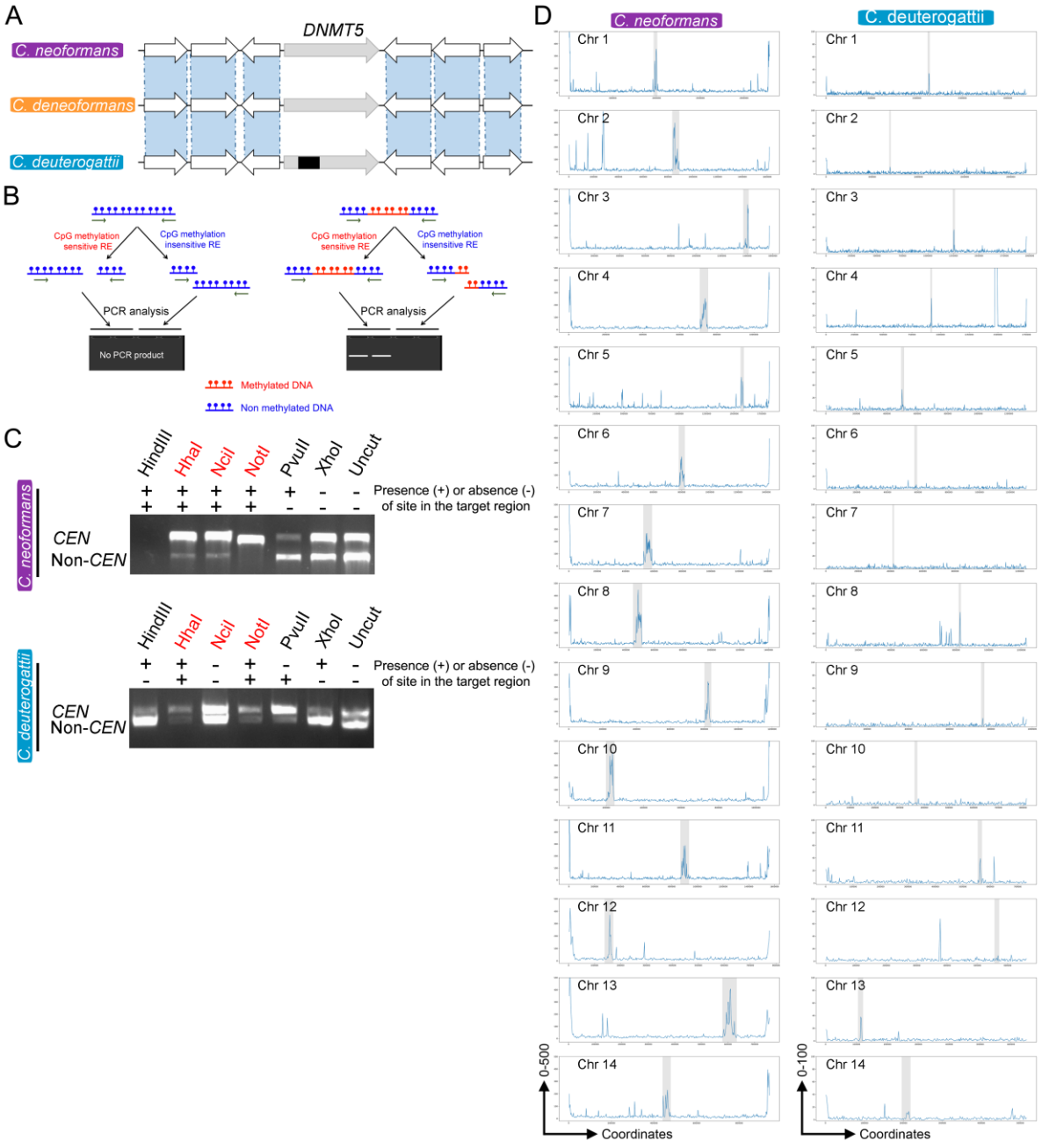
Supplementary figure S4



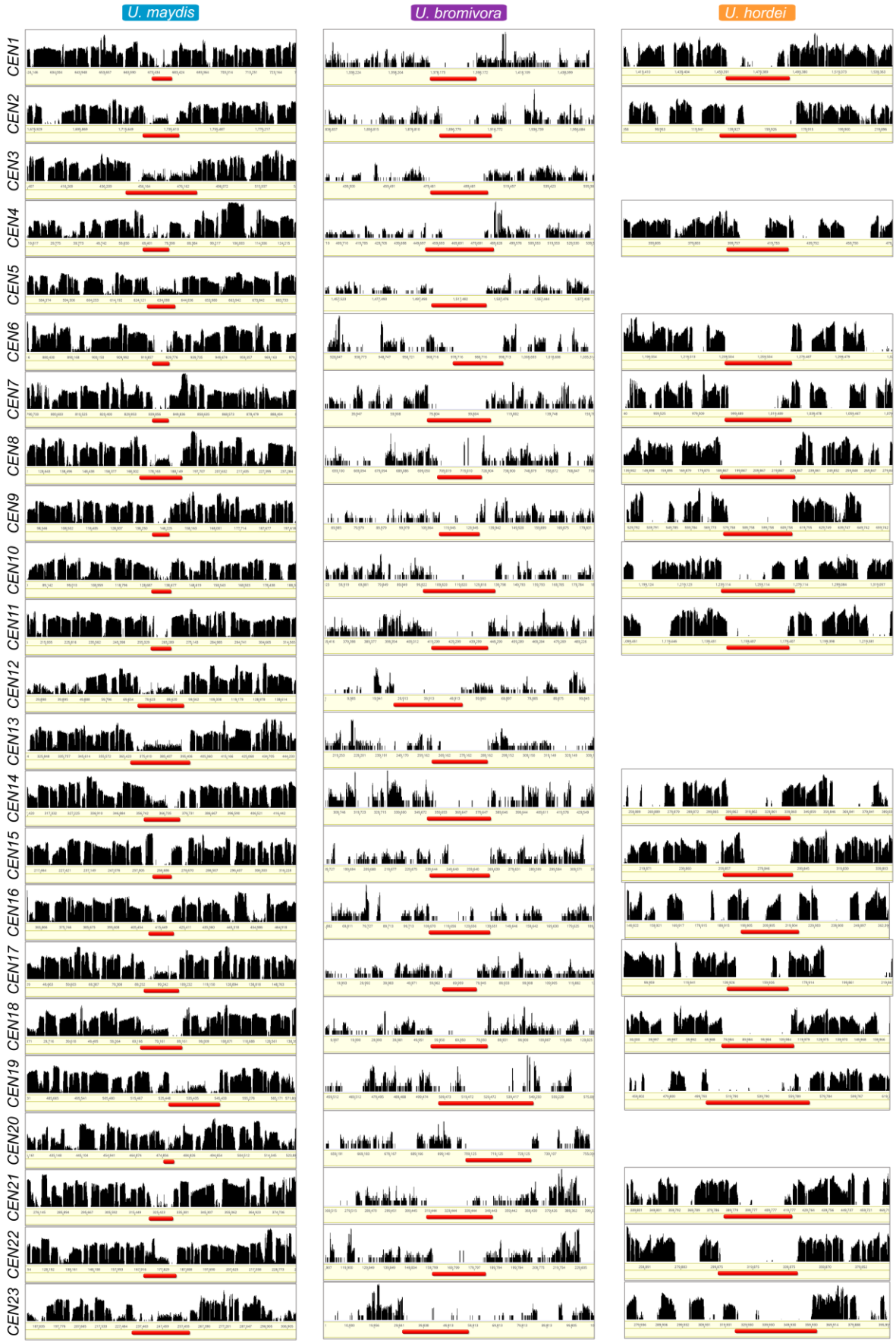
Supplementary figure S5



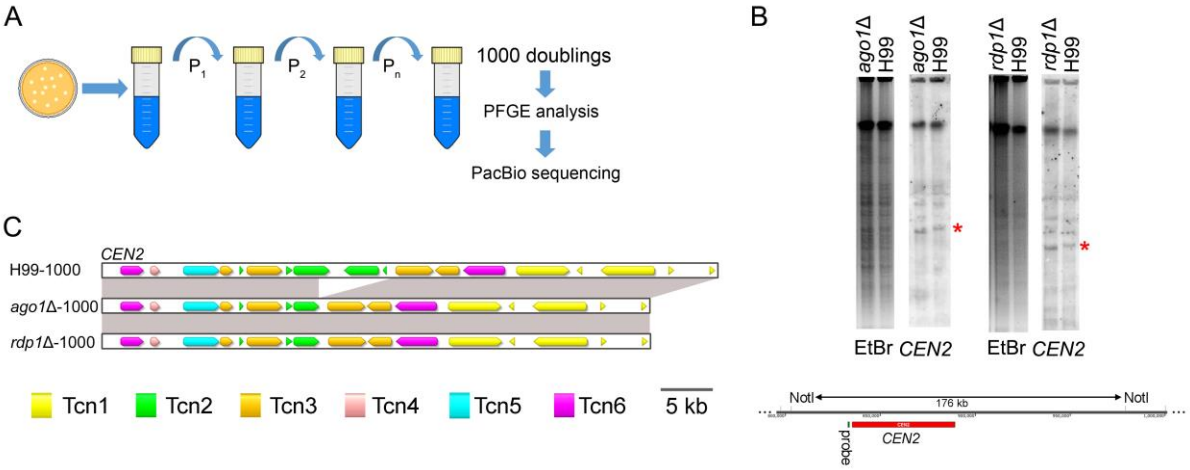
Supplementary figure S6



Supplementary figure S7



Supplementary figure S8



Supplementary references

1. Kozubowski L, *et al.* (2013) Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast *Cryptococcus neoformans*. *mBio* 4(5):e00614-00613.
2. Kozubowski L & Heitman J (2010) Septins enforce morphogenetic events during sexual reproduction and contribute to virulence of *Cryptococcus neoformans*. *Mol Microbiol* 75(3):658-675.
3. Davidson RC, *et al.* (2000) Gene disruption by biolistic transformation in serotype D strains of *Cryptococcus neoformans*. *Fungal Genet Biol* 29(1):38-48.
4. Janbon G, *et al.* (2014) Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. *PLoS Genet* 10(4):e1004261.
5. Boetzer M & Pirovano W (2014) SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* 15(1):211.
6. Thakur J & Sanyal K (2013) Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function at native physical chromosomal loci in *Candida albicans*. *Genome Res* 23:638-652.
7. Kearse M, *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.
8. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
9. Li H, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
10. Zhang Y, *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.
11. Huff JT & Zilberman D (2014) Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* 156(6):1286-1297.
12. Krueger F & Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27(11):1571-1572.
13. Robinson JT, *et al.* (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24-26.
14. Thorvaldsdottir H, Robinson JT, & Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178-192.
15. Wang X, *et al.* (2010) Sex-induced silencing defends the genome of *Cryptococcus neoformans* via RNAi. *Genes Dev* 24(22):2566-2582.
16. Ruff JA, Lodge JK, & Baker LG (2009) Three galactose inducible promoters for use in *C. neoformans* var. *grubii*. *Fungal Genet Biol* 46(1):9-16.
17. Goodwin TJ & Poulter RT (2001) The diversity of retrotransposons in the yeast *Cryptococcus neoformans*. *Yeast* 18(9):865-880.
18. Tamura K, Stecher G, Peterson D, Filipski A, & Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30(12):2725-2729.
19. Tamura K & Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10(3):512-526.

20. Soderlund C, Nelson W, Shoemaker A, & Paterson A (2006) SyMAP: A system for discovering and viewing syntenic regions of FPC maps. *Genome Res* 16(9):1159-1168.
21. Stajich JE, *et al.* (2012) FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res* 40(Database issue):D675-681.
22. Findley K, *et al.* (2012) Discovery of a modified tetrapolar sexual cycle in *Cryptococcus amylolentus* and the evolution of *MAT* in the *Cryptococcus* species complex. *PLoS Genet.* 8(2):e1002528.
23. Kamper J, *et al.* (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444(7115):97-101.
24. Kellner N, Heimel K, Obhof T, Finkernagel F, & Kamper J (2014) The SPF27 homologue Num1 connects splicing and kinesin 1-dependent cytoplasmic trafficking in *Ustilago maydis*. *PLoS Genet* 10(1):e1004046.
25. Rabe F, *et al.* (2016) A complete toolset for the study of *Ustilago bromivora* and *Brachypodium* sp. as a fungal-temperate grass pathosystem. *eLife* 5:e20522.
26. Linning R, *et al.* (2004) Marker-based cloning of the region containing the UhAvr1 avirulence gene from the basidiomycete barley pathogen *Ustilago hordei*. *Genetics* 166(1):99-111.
27. Bakkeren G, *et al.* (2006) Mating factor linkage and genome evolution in basidiomycetous pathogens of cereals. *Fungal Genet Biol* 43(9):655-666.
28. Dumesic PA, *et al.* (2015) Product binding enforces the genomic specificity of a yeast polycomb repressive complex. *Cell* 160(1-2):204-218.
29. Perfect JR, Ketabchi N, Cox GM, Ingram CW, & Beiser CL (1993) Karyotyping of *Cryptococcus neoformans* as an epidemiological tool. *J Clin Microbiol* 31(12):3305-3309.
30. Nielsen K, *et al.* (2003) Sexual cycle of *Cryptococcus neoformans* var. *grubii* and virulence of congenic a and alpha isolates. *Infect Immun* 71(9):4831-4841.
31. Heitman J, Allen B, Alspaugh JA, & Kwon-Chung KJ (1999) On the origins of congenic *MAT*alpha and *MAT*a strains of the pathogenic yeast *Cryptococcus neoformans*. *Fungal Genet Biol* 28(1):1-5.
32. Kidd SE, *et al.* (2004) A rare genotype of *Cryptococcus gattii* caused the cryptococcosis outbreak on Vancouver Island (British Columbia, Canada). *Proc Natl Acad Sci U S A* 101(49):17258-17263.